

Artificiell intelligens data – att dela eller inte dela?

PER-OLOF BJUGGREN
OCH VICKY LONG

INLÄGG Artificiell intelligens (AI) bygger precis som för människor på kunskap. Vi lär oss genom att ta del av de erfarenheter som gjorts genom historien. Vi står på jättars axlar brukar man säga. I utbildning genom skolgång får vi genom kurslitteratur ta del av de erfarenheter som tidigare gjorts. Vi lär oss också på ett mer praktiskt plan genom att tillämpa vad vi lärt och se vad som fungerar. Förenklat kan det sägas att vi lär oss hantera nya situationer genom att använda historiska data som input. På samma sätt fungerar den nya teknologin maskininläring (ML) och AI. ML och AI är bl a viktiga eftersom de ger företag en bild av trender i kundbeteende och affärsmönster, samt stödjer utvecklingen av nya produkter.

För att AI och ML ska fungera behövs s k *big data*. *Big data* är en sammanfattande term för svårhanterliga datavolymer (med traditionella metoder). Konceptet *big data* introducerades i början av 2000-talet av branschanalytikern Doug Laney (2001). Han formulerade den numera vanliga definitionen av *big data* som tre V (*volume*/stora volymer, *velocity*/ständig förändring och *variety*/stor variation). *Big data* – fångad av sensorer och robotar – är råvaran för AI och kan vara i strukturerad och ostrukturerad form. Exempel på ostrukturerade data är textfiler, bilder, videor, e-postmeddelanden. Ostrukturerade data – oorganiserad och i sin ursprungliga form – omfattar 95 procent av *big data* enligt Gandomi och Haider (2015). Strukturerade data däremot följer en fördefinierad logik som liknar ett kalkylblad där varje kolumn är märkt med

etikett. AI kan i många sammanhang processa stora mängder komplexa data effektivare än vi människor. Enligt Herbert Simon (1957) har den mänskliga hjärnan en begränsad rationalitet och det har den i större utsträckning än AI. Om så inte vore fallet skulle ett deterministiskt spel som schack vara lika enkelt för människor som det nu är för robotar.

Data är alltså den viktigaste insatsvaran för AI. Utan tillgång till data fungerar inte ML och AI. Helst ska data delas och koordineras mellan olika aktörer. Det finns fler argument som talar för datakoordinering. Rådata har karaktären av en kollektiv vara (dvs kännetecknas av icke-rivalitet i bemärkelsen att den kan konsumeras av många samtidigt). Med andra ord minskar en persons användning av data inte en annan persons användning av den (Varian 2019). Sett i det perspektivet skulle man kunna hävda att alla borde ha obegränsad tillgång till allas data. Ett annat kännetecken för AI är att det rör sig om sammanhängande värdeskapande processer mellan flera aktörer. Många företag är inte bara producenter av data, utan är också samtidigt beroende av tillgång till andras data. I de flesta fall kan värdet av data bara realiseras om alla parter samarbetar. Exempelvis krävs samarbete mellan aktörer för tillgång till den data som behövs för att självkörande bilar ska fungera på ett säkert sätt.

Det finns dessutom en tydlig fördel med att aggregera olika sorters data för att åstadkomma sambruksfördelar. Det vill säga, värdet på aggregerade data av olika karaktär är större än summan av värdena för varje typ av data. Till exempel blir en smart stad smartare när data över sopor, trafikdata, demografiska data (som exempelvis indikerar sjukvårdsbehov) och offentliga anläggningsdata slås samman för att skraddarsy samhällsmedborgares olika behov. Vidare finns det kostnadsbesparingar av att kombinera olika aktörers data i insamling,

Per-Olof Bjuggren är professor emeritus i institutionell ekonomi vid Jönköpings universitet samt associerad forskare vid Ratio – Näringslivets forskningsinstitut. p-o.bjuggren@ratio.se

Vicky Long är universitetslektor i *innovation management* vid högskolan i Halmstad och associerad forskare vid Ratio – Näringslivets forskningsinstitut. vicky.long@hh.se; vicky.long@ratio.se

sortering, lagring och överföring av data samt träning av algoritmer, vilket talar för ett fruktbart resultat av samarbete i anskaffande av data.

Dock är det flera svårigheter som måste övervinnas för att åstadkomma ett datasamarbete. Är persondata viktigt för AI är krav på personlig integritet en viktig faktor. Men även för data om affärstransaktioner som företag har samlat in och finansierat (affärsdata; B2B/*business-to-business data*), finns det svårigheter. Ska denna data vara tillgänglig för andra aktörer utan att ömsesidigt fördelaktiga avtal ingåtts? Ytterligare en kategori är offentliga data som stadskartor och annat som är viktigt för många i samhället. Om en privat aktör spenderar resurser för att utöka offentliga data, hur ska det belönas? Vi har redan (små och medelstora) företag som Univrses,¹ som sammanför smarta avfallsdata med stadskartans data, nämligen sensorinsamlade data om plats, fyllningsnivå och t o m typ av sopor för varje soptunna i Stockholm stad. Denna data kommuniceras sedan till en molnbaserad dataanalysplattform som hjälper till att ta fram optimala sophämtningsvägar, tömningsfrekvens och planering av fordonslast. Framtagning av sådan data är värdeskapande.

Givet att tillgång till data är vad som krävs för att AI ska utvecklas kommer incitament och regler för anskaffning och handel i fokus. Samarbete genom avtal måste lösas avseende kontroll och rättigheter. Det finns en stor oklarhet om vilket institutionellt ramverk i form av lagar som bör styra data och dess användning liksom hur rätten till den *output* som produceras i form av ny kunskap ska se ut. Maskingenererad kunskap hamnar mellan stolarna i lagstiftningen. Det finns i princip ingen möjlighet att få en tydlig exklusiv äganderätt till den data som samlas in och produceras av

robotar och maskiner. Upphovsrätten skyddar bara mänsklig kreativitet. Data-basdirektivet 1996 (*sui generis*), som bara finns i EU:s lagstiftning, erkänner visserligen en exklusiv rätt till en databas när det finns en betydande investering. Det är dock ännu inte klarlagt om den erforderliga investeringen avser maskingenererade data eller av människor insamlade data. Det står heller inte om det avser delar av eller en hel databas. Detta påverkar naturligtvis incitamenten då äganderätt är en grundsten i en marknadsekonomi med priser som allokeringmekanism och vinstdrivande konkurrerande företag.

Bristen på *de jure* äganderätt har ersatts av en *de facto* teknisk kontroll av data genom smartcards, PINs, biometrisk identifikation och bilaterala kontrakt. Sådan *de facto* kontroll med hjälp av bilaterala kontrakt begränsar möjligheten till att skala upp koordinering av data mellan många olika aktörer. Den koordinering som krävs för att ML ska fungera så bra som möjligt hämmas. Återanvändning av data påverkas således negativt. För exempelvis affärsdata ("B2B data") har det observerats att det inte finns tillräckligt med datadelning och mycket data förblir låst (av t ex plattformägare) och är inte tillgänglig för återanvändning (Crémer m fl 2019).

Den nya ekonomin med AI och ML har implikationer för företags maktställning, strategi och personlig integritet. Det är i dessa frågor som AI hamnar i en ny terräng. Vi pekade på likheter mellan människor och AI när det gäller inlärning men resultaten av inlärning kan ge mycket mer makt hos AI än hos människor. Det finns stordriftsfördelar både på produktionssidan och efterfrågesidan som gör att en producent av AI-genererade nyttigheter kan bli helt dominerande. Stordriftsfördelarna på produktionssidan består av möjlighe-

¹ <https://univrses.com/>.

terna att fördela kostnader för insamling av data och träning av algoritmer på en större volym producerade tjänster. De flesta större aktörer, inte minst plattformsföretag som Tesla eller Amazon, kan anses ha stordriftsfördelar på produktionssidan: ju fler kunder de har, desto lägre driftskostnad per kund för att t ex upprätthålla en molntjänst. Efterfrågefördelarna kan illustreras med Facebook, en plattform som blir attraktivare ju fler användare som finns och kan ingå i kretsen vänner. Andra exempel på sambruksfördelar som gör att en plattformsaktör blir dominerande är: LinkedIn som kan berätta för dig som individ om jobb du kan vara intresserad av (eftersom de samlar olika jobbanonseringsdata); Amazon som berättar för Semper (barnmat) att kunden som köpte blöjor från Pampers kan vara intresserad av Sempers välling; en (geo-)sökalgorithm som säger till att din favorit Max-hamburgare bara ligger fem minuters promenad bort.

Vid användning av gratis dataplattformar, som exempelvis Google eller Apoteket, genereras en biprodukt i form av personlig information som är värdefull i marknadsföring och kan säljas. Sådan handel med persondata ("B2P marketing") är mycket omtvistad inte minst med tanke på personlig integritet. När personuppgifterna väl samlats in har individen liten kunskap eller kontroll av hur företaget, som samlade in uppgifterna, senare kommer att använda dessa uppgifter. Företaget kan sälja konsumentens data till tredje part med vinst, som konsumenten inte får någon del av. Individer kan åsamkas extra kostnader när en tredje parts användning av personuppgifter resulterar i att de t ex får skräppost och utsätts för negativ prisdiskriminering. Alltmer strikta regler om att värna om personlig integritet har tagits fram som de i AI-sektorn måste förhålla sig till. Dock är persondata viktiga i andra sammanhang än marknadsföring.

För utveckling av t ex självkörande bilar är det viktigt att få feedback från persondata. Platsrelaterad, beteendemässiga data och "bulk PI" (personligt identifierbar information) är värdefulla i detta sammanhang. Det sker därför en intensiv utveckling, benämnd dataekologi, med (*blockchain*-baserad) identitetsdöljande teknik som gör det möjligt att använda persondata på ett etiskt acceptabelt sätt (se Hardjono m fl 2019). I ett *blockchain*-system kan en teknisk nod kryptografiskt bevisa vem som är medlem utan att avslöja medlemmens identitet. Syftet är att hitta en balans mellan att använda och skydda individers data, dvs en balans mellan den effektivitetspotential som finns av att ha tillgång till personuppgifter och människors medvetenhet, känslighet och önskan att skydda personuppgifter. Transparens, kontroll och integritet eftersträvas.

Med tanke på de många stordriftsfördelar som finns inom AI är också konkurrenslagstiftningen ett hett ämne som visats i rättsfall för stora företag med dominerande marknadsställning inom den nya ekonomin. Nuvarande konkurrenslagstiftning, i synnerhet dess "Essential Facility Doctrine (EFD)" om leveransvägran, bygger på idén att ett monopolistiskt företag har en skyldighet att dela sina faciliteter (inklusive data) med alla som ber om tillgång (inklusive konkurrenter). Detta är dock bara relaterat till en monopolistisk situation baserad på immateriell äganderätt, snarare än allmänt konkurrensbegränsande beteende. Att använda befintligt lagverk på data innebär stora svårigheter (Colangelo och Maggolino 2017). Juridiken tycks inte heller inom detta område ha hunnit med att finna ett ramverk för AI att verka inom.

Sammanfattningsvis finns det en problematik i åtkomsten av data som är den viktigaste insatsvaran för AI. Det rättsliga ramverket för AI behöver ses över. Äganderätter finns inte på samma

sätt som inom andra områden. Ensamt rätt etableras genom bilaterala kontrakt och tekniska lösningar som hindrar vidare-/återanvändning av data. Men för att ytterligare utveckla teknologin behövs tillgång till persondata, vilket kan vara kränkande för den personliga integriteten och användas för syften som är skadliga för samhället. Skandalen Cambridge Analytica är ett exempel. Där samlade det brittiska konsultföretaget Cambridge Analytica in personuppgifter som tillhörde miljontals Facebook-användare (utan deras medgivande). Dessa data visade sig vara användbara för politisk påverkan i exempelvis presidentvalskampanjer. Det finns därför strikta regler om användning av persondata. För att komma åt problem med integritet pågår en utveckling av ny teknologi och nya affärsmetoder som tillgodoser både problemet med integritet och behov av åtkomst av persondata för utveckling av samhällsnyttiga AI-tjänster.

Den ekonomiska debatten i Sverige handlar inte så mycket om AI. Det skrivs sällan debattartiklar som handlar om AI, vilket är anmärkningsvärt med tanke på hur den nya teknologin påverkar vår

ekonomi. I börshandeln under det senaste årtiondet har bolag som anammade den nya teknologin uppmärksammas och genererat stora vinster för placerare. Vad som styr AI och betydelsen av data borde vara mer i fokus i den ekonomiska debatten.

REFERENSER

Colangelo, G och M Maggolino (2017), "Big Data as a Misleading Facility", *European Competition Journal*, vol 13, nr 2, s 249–281.

Crémer, J, Y-A de Montjoye och H Schweitzer (2019), "Competition Policy for the Digital Era", Report for the European Commission.

Gandomi, A och M Haider (2015), "Beyond the Hype: Big Data Concepts, Methods and Analytics", *International Journal of Information Management*, vol 35, s 137–144.

Hardjono T, D L Shrier och A Pentland (2019) (red), *Trusted Data: A New Framework for Identity and Data Sharing*, MIT Press, Cambridge MA.

Laney, D (2001), "3D Data Management: Controlling Data Volume, Velocity and Variety", META Group Research Note 6.

Simon, H A (1957), *Models of Man*, John Wiley, New York.

Varian, H (2019), "Artificial Intelligence, Economics, and Industrial Organization", i Agrawal, A, J Gans och A Goldfarb (red), *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, Chicago.